Region of Interest Video Coding for Ultrasound Imagery

Brian C.S. Loh & Patrick H.H. Then

Abstract—**Echocardiograms are the result of utilizing ultrasound devices to capture cardiac structures and dynamics. The results of these scans, which are depicted as video or images, are assessed by physicians to determine heart conditions. Although ultrasound devices are making way into diverse medical scenarios, their utility has been hampered by the need for extensive operator training to perform examinations, and medical experience required for analysing imagery. Telesonography or tele-ultrasound, allows for the guidance and interpretation of ultrasound acquisitions by distant medical experts. Since video content sharing necessitates large amounts of bandwidth, encoding techniques can be applied beforehand to reduce the size of exchanged data. This paper introduces a healthcare solution that aims to support the encoding and compression of medical imagery. The proposed application combines several image processing techniques with ROI video coding algorithms to achieve optimal size and quality trade-offs. Experiments show that file size reductions from 26% to 55% can be achieved when encoding with the application.**

Keywords—*v*ideo coding, region of interest, telesonography, echocardiogram

I. Introduction

A. *Cardiac Ultrasound*

Over the decades, ultrasound has remained a vital imaging modality in cardiology. Compared with other technologies such as computed tomography, nuclear medicine, and magnetic resonance imaging, ultrasound provides several benefits including low costs, sustainability, versatility, and safety [1]. An echocardiogram video is the result of utilizing ultrasound to capture cardiac structures and its dynamics [2], [3]. A complete evaluation involves positioning a scanning device at various locations on a patients' chest, and then recording sequences representing different cross sections of the heart. The results of these scans, which are depicted as video or images, are assessed by physicians to determine cardiac conditions. Advances in technology have enabled the

Brian C.S. Loh

Faculty of Engineering, Computing and Science
Swinburne University of Technology, Sarawak Campus
Kuching, Malaysia

Patrick H.H. Then

Faculty of Engineering, Computing and Science
Swinburne University of Technology, Sarawak Campus
Kuching, Malaysia

miniaturization of ultrasound systems which in turn has increased the usage of portable ultrasound devices in primary care and nonhospital settings.

B. *Telesongraphy*

Although ultrasound devices are making way into diverse medical scenarios, their utility has been hampered by the need for extensive operator training to perform examinations, and medical experience required for analysing imagery [4], [5]. Without trained personnel capable of conducting accurate interpretations, the potential for misapplication or incorrect diagnosis may hinder its adoption [6]. Training all potential users in ultrasound device usage would be an extremely challenging task [7]. Instead, novice sonographers with limited experience can obtain remote assistance from expert advisors. Telesonography or tele-ultrasound, allows for the guidance and interpretation of ultrasound acquisitions by distant medical experts [8], [9]. Its main objective is the sharing and exchange of ultrasound imagery for the purpose of obtaining remote diagnosis, either in real time or following a time delay. By utilizing telesonography, important choices can be made, such as the selection of destination hospital, the decision to evacuate a patient in a wilderness setting, or the choice to use scarce resources in underdeveloped regions [10].

C. *Video Coding*

At the remote location, medical videos obtained using portable equipment require transmission to experts for diagnosis. Since video content sharing necessitates large amounts of bandwidth, encoding techniques are applied beforehand to reduce the size of exchanged data [11]. Medical video transmission systems are reliant upon two significant components: video compression technology and the wireless networks utilized for communication [12]. Video coding refers to the process of compressing input signal data by reducing redundancies inherent in image and video, such as spatial, temporal, statistical and psychovisual information [13].

H.264/AVC is the current state-of-the-art compression standard that provides efficient encoding capable of matching the underlying transmission medium data rate, and a low complexity decoder that allows for power saving implementations on mobile devices [14]. In spite of the growth in communication technologies and video coding standards, telesonography systems still face bandwidth and quality issues when deployed in low bandwidth conditions. Due to slow transfer speeds, high data compression rates are necessary for timely exchange of critical information. This in turn causes the degradation of visual elements which negatively affects diagnostic quality of medical videos.

D. *Region of Interest Video Coding*

A method for dealing with this issue is region of interest (ROI) video coding which seeks to improve the visual characteristics of crucial portions in a frame. A frame can typically be divided into multiple segments, with the most important part, the ROI, being the main focus of observing audiences. Therefore, the goal of ROI video coding is to encode all objects within the relevant sections with a higher quality as compared to its background. The benefit of this approach is an improved viewing experience as the ROI can be more accurately examined. Moreover, the amount of data required to send the video file can be reduced if the ROI area is smaller than its surroundings.

Medical imagery usually comprise of critical diagnosis zones and surrounding regions which provide additional contextual information. In these cases, choosing the ROI is a necessary task which can be done either explicitly or implicitly [15]. Explicitly defined ROIs involve user input such as mouse click selection, finger presses or eye movement tracking, whereas implicitly defined ROIs rely on content recognition algorithms that automatically detect objects based on specified conditions. An advantage of user selected ROIs is that they are precisely chosen depending on the needs of the viewer. On the other hand, if a video contains multiple ROIs which are in constant motion, automated detection methods can reduce the time needed for locating and delineating these regions.

E. *Contribution*

Medical videos are generally stored in high quality formats to allow for complete and accurate diagnosis. Because of this requirement, files commonly occupy large portions of disk space. Though essential in the decision making process, the effectiveness of these videos would be limited in areas or conditions with slow internet connectivity. Due to enormous data volumes, lengthy transfer durations may hinder response times. Additionally, although existing compression methods can reduce file sizes, their effect on video quality may impede detailed examinations. To overcome these issues, this paper introduces a healthcare solution that aims to support the encoding and compression of medical imagery. The proposed application combines several image processing techniques with ROI video coding algorithms to achieve optimal size and quality trade-offs.

II. Literature Review

Some researchers have focused on user defined ROIs [16], [17] while others employed automated methods [14], [18]. Rao et al. [16] proposed an elastic ROI coding algorithm for mobile medical assessments of acute childhood respiratory distress. Physicians could specify an ROI which varied from 25% to 50% of the total frame. This ROI would then be encoded at diagnostic quality levels previously obtained through expert feedback. Khire et al. [17] followed a similar approach for the purpose of surgical telementoring. Remotely viewing specialists were given the ability to choose an ROI location on a frame which was divided into nine equal segments. Furthermore, visual quality of the selected region could be manually controlled to achieve diagnostically lossless standards. Panayides et al. [14], [18] decided on a different technique for ROI selection. A pixel-based segmentation algorithm was used to automatically estimate motions of plaque and wall boundaries in ultrasound videos. From there, ROI areas and background were encoded at contrasting quality levels to reduce bandwidth requirements while maintaining adequate diagnostic conditions. Compared to the above works, the proposed application takes a hybrid approach for ROI selection by allowing users to manually delineate important areas, and then dynamically estimating the necessary visual quality for those zones. Instead of enforcing fixed quality values for background and ROI, block level processing is performed on each video frame. This ensures that the encoding process takes into account not only the user specified ROI, but the frame contents as well.
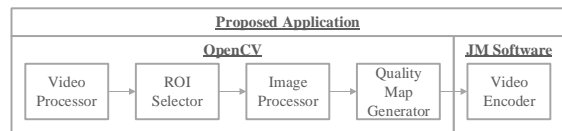
III. Framework

A. *Overview*

Figure 1. System modules

Figure 1 presents the system overview which consists of five modules: video processor, ROI selector, image processor, quality map generator and video encoder. The proposed application was built on the C++ programming language in combination with the Open Source Computer Vision Library v2.4.11 (OpenCV) and the H.264/AVC JM Reference Software v19.0. OpenCV was selected for its cross-platform compatibility in addition to its advanced video and image processing functions. For video encoding, the JM Reference Software was chosen as it contained the necessary standards and features for creating H.264/AVC outputs.
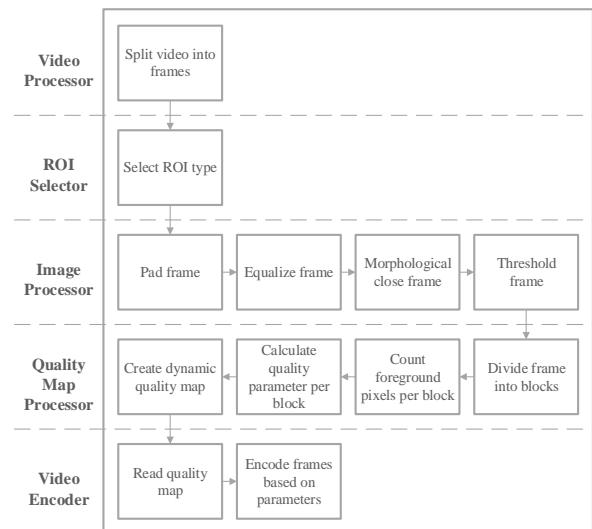

Figure 2. Application process flow

The processing phase incorporates all modules and consists of 12 steps which are detailed in Figure 2. The process flow starts from the input of an echocardiogram video to the output of a compressed version.

1) Splitting a video into frames – The number of frames in a video are determined, and each frame is saved as an image for further processing.

2) Selection of ROI type – The first video frame is displayed to the user who is given the choice of selecting a single or multiple ROIs. The user can draw rectangular sections over important zones in the frame which would later be encoded with a higher quality level than background areas.

3) Padding frames – If a frame cannot be divisible into full blocks of 16 X 16 pixels, edges are padded to reach the appropriate dimensions.

4) Equalizing frames – Certain ultrasound images may have low contrast levels which can affect the thresholding process. Because of this, frames are equalized to ensure proper distribution of intensities.

5) Morphological closing frames – Speckle noise which is usually present in ultrasound images can interfere with the segmentation procedure. To alleviate this issue, a morphological close is performed to reduce the amount of noise in an image.

6) Thresholding frames – To prepare an image for segmentation, it is converted to black and white based on an appropriate threshold level chosen by Otsu's method [19].

7) Dividing frame into blocks – A frame is divided into blocks of 16 X 16 pixels. These blocks represent macroblocks, which are the basic processing units for H.264/AVC encoders.

8) Counting foreground pixels in each block – The number of white pixels in a block are counted to determine the amount of information present. Since thresholding was performed, foreground objects would appear in white while the background remains black.

9) Calculating quality parameter for each block – A quality parameter is calculated based on the information present in a block. This value equates to the quantization parameter (QP) used in H.264/AVC encoders.

10) Creating a dynamic quality map – All quality values for each block in every frame are written into a text file.

11) Reading quality map as input – The encoder reads and stores all values from the quality map file.

12) Encoding frames based on quality map parameters – Each block is encoded based on its respective quality level.

B. *Image and Quality Map Processor*

Both the image and quality map processor are prime components in the proposed application. They consist of major functions which directly control the amount of compression and visual quality maintained in the final output. Figure 3 demonstrates the processing done by the modules on a sample ultrasound frame. The first image is the initial frame of an acquired echocardiogram video. In the subsequent image, equalization is performed to enhance overall contrast, hence improving clarity of cardiac entities. After that, a morphological close operation is initiated to remove speckle noise, followed by thresholding of the frame to separate background and foreground objects. Next, the frame is divided into blocks which will be processed individually. The quality of a block is calculated based on the amount of pixels present within the block. If a block contains mostly white pixels, it is assumed to possess a high degree of information. The final frame shows the quality map overlaid on the thresholded image. The numbers on each block correspond to QP values that will be used to encode the block. The range of QP starts from 0 (lossless quality) to 51 (worst quality). By default, the JM Reference Software sets QP to 28 for encoding. Because of that, fully filled blocks are initialized with a value of 28. As less information is present in a block, a higher level of compression is applied. Consequently, for every 10% reduction of white pixels, QP is increased by 1. Lastly, background blocks completely filled with black pixels are assigned a QP of 42. Although the value of 51 could provide a higher level of compression, it was not chosen because visual artefacts appeared when QP was above 42.
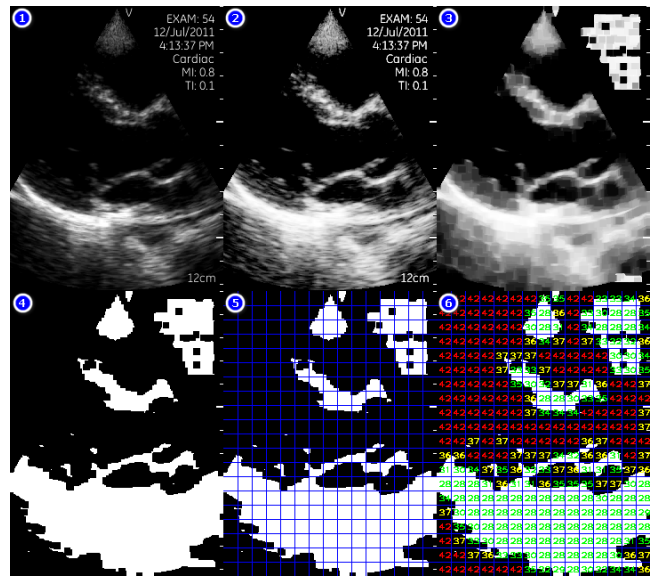

Figure 3. Sample ultrasound frame

The formula for block quality calculation is described in Table 1. Once all blocks in the current frame are processed, the same procedure is repeated on the remaining frames. This enables dynamic frame quality adjustments based on block level information, thus ensuring optimal file size reduction and visual quality preservation.

TABLE 1. BLOCK QUALITY CALCULATION

Variable	Value
Initial QP	28
Offset	9
Block Coverage	Total white pixels $\div$ $(16 \times 16) \times 100$
Formula	
Initial QP + (Offset – [[(Offset $\times$ Block Coverage $\div$ 100)]])	

IV. Experiment and Results

Experiments were performed with the goal of determining file size reduction after quality map encoding in both no ROI and single ROI scenarios. Furthermore, objective metrics were used to estimate the amount of visual information preserved in an encoded video. A parasternal long axis scan obtained from a portable ultrasound device was used as input. The video consisted of 15 frames in total and was encoded with the JM Reference Software baseline profile at 28 QP. Eight experiments with variations in image processing settings were run. Experiment parameters are shown in Table 2 and results displayed in Figure 4.

TABLE 2. EXPERIMENT PARAMETERS

Exp.	Equalization	Morphological Close	Threshold	ROI
1	Yes	Yes	Yes	No
2	Yes	No	Yes	No
3	No	Yes	Yes	No
4	No	No	Yes	No
5	Yes	Yes	Yes	Yes
6	Yes	No	Yes	Yes
7	No	Yes	Yes	Yes
8	No	No	Yes	Yes

In each experiment, the effects of equalization and morphological close were tested. Considering that equalization improves image contrast, it can be expected that more pixels are detected. This would in turn cause the quality map processor to assign lower QP values to the blocks, and therefore increase overall visual quality. A similar outcome occurs when morphological close is performed. Since holes in an image are filled with white pixels, certain blocks would be given better quality values.
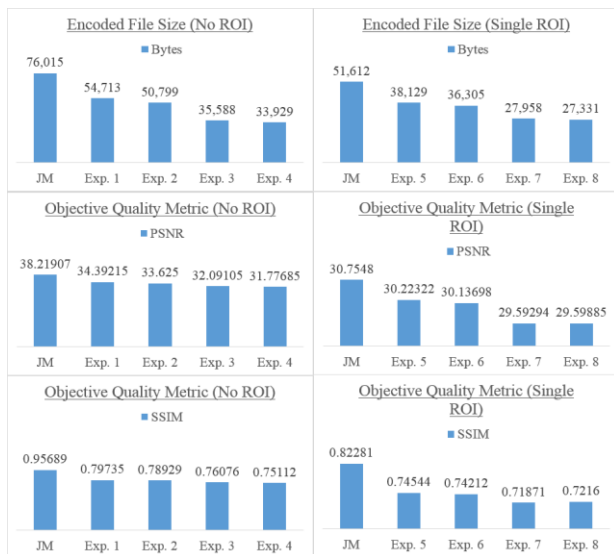

Figure 4. Experiment results

Results show file size and objective quality metric comparisons between the four experiments and the encoded ultrasound video. Depending on experiment parameters, file size reductions from 26% to 55% were achieved when encoding with the quality map. As anticipated, disabling equalization and morphological close improves file compression. Two widely adopted objective metrics, peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) were used to determine visual quality of the encoded videos. PSNR measures the amount of noise introduced after compression while SSIM estimates similarity by means of luminance, contrast and structural comparisons. There exists a strong correlation between file size and objective quality as visual details are degraded during the compression process. However, objective quality metrics do not necessarily determine the presence or absence of utility in a video. Take for instance the experiment outputs displayed in Figure 5. Although the objective metrics show reduced quality, visual comparison of the frames do not indicate significant degradation apart from slight emergence of artefacts in darker regions.
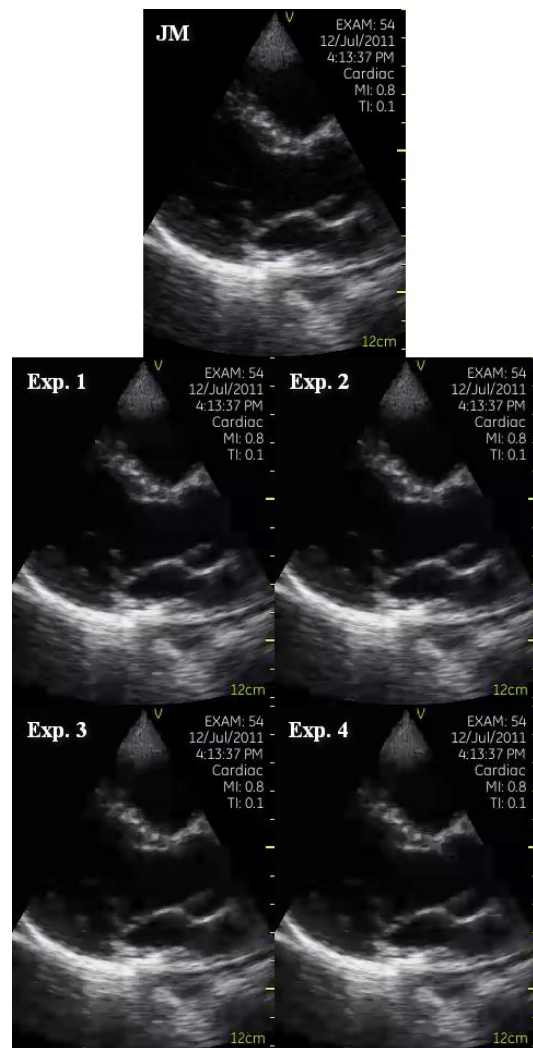

Figure 5. Experiment outputs (no ROI)

v. **Conclusion**

Medical imagery are essential for diagnosis and decision making purposes. As a consequence, the preservation of visual quality is of the upmost importance for ensuring precise judgements. Moreover, the reduction of file size is necessary to provide rapid exchange of critical information. The proposed application addresses both these needs through the implementation of image processing and ROI video coding. Future works aim to include other views of the heart as well as alternative imaging modalities such as magnetic resonance imaging and computed tomography. Additionally, apart from purely focusing on block pixel information, the motion of objects over a collection of frames will be studied.

References

[1] P. A. Gladding, A. Cave, M. Zareian, K. Smith, J. Hussan, P. Hunter, F. Erogbogbo, Z. Aguilar, D. S. Martin, and E. Chan, "Open access integrated therapeutic and diagnostic platforms for personalized cardiovascular medicine," Journal of personalized medicine, vol. 3, no. 3, pp. 203-237, 2013.

[2] S. Ebadollahi, S.-F. Chang, and H. Wu, "Echocardiogram videos: summarization, temporal segmentation and browsing." pp. I-613-I-616 vol. 1.

[3] S. Ebadollahi, S.-F. Chang, and H. Wu, "Automatic view recognition in echocardiogram videos using parts-based representation." pp. II-2-II-9 Vol. 2.

[4] V. Popov, D. Popov, I. Kacar, and R. D. Harris, "The feasibility of real-time transmission of sonographic images from a remote location over low-bandwidth Internet links: a pilot study," American Journal of Roentgenology, vol. 188, no. 3, pp. W219-W222, 2007.

[5] A. Parsai, I. Zerizer, J. Hohmann, G. Bongartz, C. Beglinger, and G. Sperandeo, "Remote sonographic interpretation: Comparison of standardized video clips to still images," Journal of Clinical Ultrasound, vol. 40, no. 8, pp. 495-501, 2012.

[6] B. G. Choi, M. Mukherjee, P. Dala, H. A. Young, C. M. Tracy, R. J. Katz, and J. F. Lewis, "Interpretation of remotely downloaded pocket-size cardiac ultrasound images on a web-enabled smartphone: validation against workstation evaluation," Journal of the American Society of Echocardiography, vol. 24, no. 12, pp. 1325-1330, 2011.

[7] J. Pyke, M. Hart, V. Popov, R. D. Harris, and S. McGrath, "A tele-ultrasound system for real-time medical imaging in resource-limited settings." pp. 3094-3097.

[8] P. McBeth, I. Crawford, C. Tiruta, Z. Xiao, G. Q. Zhu, M. Shuster, L. Sewell, N. Panebianco, D. Lautner, and S. Nicolaou, "Help is in your pocket: the potential accuracy of smartphone-and laptop-based remotely guided resuscitative telesonography," TELEMEDICINE and e-HEALTH, vol. 19, no. 12, pp. 924-930, 2013.

[9] L. Pian, L. M. Gillman, P. B. McBeth, Z. Xiao, C. G. Ball, M. Blaivas, D. R. Hamilton, and A. W. Kirkpatrick, "Potential use of remote telesonography as a transformational technology in underresourced and/or remote settings," Emergency medicine international, vol. 2013, 2013.

[10] A. S. Liteplo, V. E. Noble, and B. H. Attwood, "Real-time video streaming of sonographic clips using domestic internet networks and free videoconferencing software," Journal of Ultrasound in Medicine, vol. 30, no. 11, pp. 1459-1466, 2011.

[11] C. J. Debono, B. W. Micallef, N. Y. Philip, A. Alinejad, R. S. Istepanian, and N. N. Amso, "Cross-layer design for optimized region of interest of ultrasound video data over mobile WiMAX," Information Technology in Biomedicine, IEEE Transactions on, vol. 16, no. 6, pp. 1007-1014, 2012.

[12] A. Panayides, M. Pattichis, C. Pattichis, and A. Pitsillides, "A tutorial for emerging wireless medical video transmission systems [Wireless Corner]," Antennas and Propagation Magazine, IEEE, vol. 53, no. 2, pp. 202-213, 2011.

[13] J. Zeng, O. C. Au, W. Dai, Y. Kong, L. Jia, and W. Zhu, "A tutorial on image/video coding standards." pp. 1-7.

[14] A. Panayides, M. S. Pattichis, C. S. Pattichis, C. Loizou, M. Pantziaris, and A. Pitsillides, "Atherosclerotic plaque ultrasound video encoding, wireless transmission, and quality assessment using H. 264," Information Technology in Biomedicine, IEEE Transactions on, vol. 15, no. 3, pp. 387-397, 2011.

[15] Y.-B. Lin, and X.-M. Zhang, "Recent developments in perceptual video coding." pp. 259-264.

[16] S. P. Rao, N. S. Jayant, M. E. Stachura, E. Astapova, and A. Pearson-Shaver, "Delivering diagnostic quality video over mobile wireless networks for telemedicine," International Journal of Telemedicine and Applications, vol. 2009, pp. 1, 2009.

[17] S. Khire, S. Robertson, N. Jayant, E. A. Wood, M. Stachura, and T. Goksel, "Region-of-interest video coding for enabling surgical telementoring in low-bandwidth scenarios." pp. 1-6.

[18] A. Panayides, Z. C. Antoniou, Y. Mylonas, M. S. Pattichis, A. Pitsillides, and C. S. Pattichis, "High-resolution, low-delay, and error-resilient medical ultrasound video communication using H. 264/AVC over mobile WiMAX networks," Biomedical and Health Informatics, IEEE Journal of, vol. 17, no. 3, pp. 619-628, 2013.

[19] N. Otsu, "A threshold selection method from gray-level histograms," Automatica, vol. 11, no. 285-296, pp. 23-27, 1975.